

Estimating financial risk using piecewise Gaussian processes

I. García^{a,1}, J. Jiménez^{b,c}

^a*Departamento de Cómputo Científico y Estadística and Centro de Estadística y Software Matemático (CESMa), Universidad Simón Bolívar, Sartenejas, Venezuela*

^b*Laboratorio de Fenómenos no Lineales, Escuela de Física, Facultad de Ciencias, Universidad Central de Venezuela*

^c*Red de Estudios Interdisciplinarios, Academia Nacional de Ciencias Físicas, Matemáticas y Naturales, Venezuela*

Abstract

We present a computational method for measuring financial risk by estimating the Value at Risk and Expected Shortfall from financial series. We have made two assumptions: First, that the predictive distributions of the values of an asset are conditioned by information on the way in which the variable evolves from similar conditions, and secondly, that the underlying random processes can be described using piecewise Gaussian processes. The performance of the method was evaluated by using it to estimate VaR and ES for a daily data series taken from the S&P500 index and applying a backtesting procedure recommended by the Basel Committee on Banking Supervision. The results indicated a satisfactory performance.

Keywords: Forecasting, Econometrics, Value at Risk, Expected Shortfall

Email address: gmosquer@usb.ve (I. García)

¹Telefax: +582129063364

1. Introduction

Since the classic study by Haavelmo (1944) on the probability approach in econometrics, there has been a tendency to assume that a series of values of an asset, $\{V_\tau; \tau = 1, \dots, t\}$, represents a random process with density $\rho_\tau(V)$. In fact, from the point of view of financial practice, what is interesting is to be able to make predictions conditioned by a particular information set. Thus, what we really want to model is the density $\rho_{t+1}(V|I_t)$, conditional on the information I_t known at t , from which the expected price of the asset:

$$\bar{V}_{t+1} = \int_0^\infty V \rho_{t+1}(V|I_t) dV$$

and the deviations in these prices

$$\sigma_{t+1} = \sqrt{\int_0^\infty (V - \bar{V}_{t+1})^2 \rho_{t+1}(V|I_t) dV}$$

can be calculated.

The expected return, $\bar{R}_{t+1} = (\bar{V}_{t+1} - V_t)/V_t$, can thus be obtained from the first integral, whilst the standard deviation is a measure of the uncertainty that is directly related to indexes commonly used in risk management, such as the Value at Risk (*VaR*) and Expected Shortfall (*ES*). The problem, of course, is that there is only one realization for each random process, thus the conditional distribution is an unobservable quantity that can only be inferred.

In this context, this study has a double aim: firstly, to use (in a way that we will describe in detail later) information about the evolution of prices from similar initial conditions, as the information I_t that conditions the distribution we are looking for, and secondly, to design an inference model based on

the assumption that the underlying random processes are piecewise Gaussian processes.

Due to the simplicity and versatility of Gaussian Processes (*GP*) for the modeling of arbitrary functions, they have become the basis for several techniques that have been developed to analyze several types of spatial and temporal problems (see e.g. Brahim-Belhouari and A. Bermak (2004), Bukkapatnam and Cheng (2010), Deisenroth et al. (2009), di Sciascio and Amicarelli (2008), Gosoniua et al. (2009), Kim et al. (2005), Ko and Fox (2009) and Wang et al. (2008)). In addition, modeling with *GP* has an advantage that makes it particularly attractive from an econometric point of view, which is that it does not only permit predictions of the values of a series to be made, but also generates predictive distributions. This means that both the volatilities (given by the standard deviations of the predictive distributions) and indexes of risk such as *VaR* can be estimated directly. Nevertheless, to the best of our knowledge, this approach has not been previously used for the estimation of financial risk indexes.

Furthermore, our *GP* based models are local in the sense that the involved parameters depend on the information I_t , that in turn changes with time. As a consequence, the conditional predictive variances are time dependent and the models are thus heteroscedastic. In contrast to the ARCH/GARCH type models, however, there is no equation that controls the evolution of the volatility (see Engle et al. (2008)), and this change is instead driven by changes in I_t .

In the following section we introduce the specific type of piecewise models that we apply to estimate the risk in a financial time series. In section III we

apply a validation procedure (backtesting) to evaluate the model according to recommendations given by the Basel Committee on Banking Supervision (Basel I (1995), Basel II (1996) and Hull (2007)). Concluding remarks are presented in Section IV.

2. Prediction of the Value at Risk and Expected Shortfall using Piecewise Gaussian Processes

In what follows, we consider the simple case where we wish to make an estimation, $\hat{\rho}_{t+1}(V|I_t)$, of the conditional distribution $\rho_{t+1}(V|I_t)$ for a known series of the values of an asset V_1, V_2, \dots, V_t , in order to predict the return one step ahead:

$$\hat{R}_{t+1} = (\hat{V}_{t+1}/V_t) - 1; \quad \hat{V}_{t+1} = \int_0^\infty V \hat{\rho}_{t+1}(V|I_t) dV \quad (1)$$

and the volatility measured by the deviation:

$$\hat{\sigma}_{t+1}^R \equiv \frac{\hat{\sigma}_{t+1}}{V_t} = \frac{1}{V_t} \sqrt{\int_0^\infty (V - \hat{V}_{t+1})^2 \hat{\rho}_{t+1}(V|I_t) dV} \quad (2)$$

Specifying the information by which it is conditioned is therefore essential. We thus assume, as for the majority of the models used in financial econometrics, that events which occurred most recently are of greater relevance when we wish to establish conditions in the near future. This is the same as assuming that I_t is contained in the sequence of values of the recent past: $V_{t-l+1}, V_{t-l+2}, \dots, V_{t-1}, V_t$, corresponding to a rolling window of size l . Part of the original contribution of our proposal has to do with the way in which I_t is extracted from this data window. In this sense, our specific interest is focused on using the values of the asset which follows patterns of behavior of the data that are similar to those observed in the current window.

In order to explain in more detail what we mean, let's start by considering all possible segments of historical information that can be constructed with l consecutive values of the prices of an asset:

$$\begin{aligned}\mathbf{v}_1 &= (V_1, V_2, \dots, V_l) \\ \mathbf{v}_2 &= (V_2, V_3, \dots, V_{l+1}) \\ &\vdots \\ \mathbf{v}_{t-l+1} &= (V_{t-l+1}, \dots, V_{t-1}, V_t)\end{aligned}$$

The behavior patterns are understood to be the standardizations:

$$\mathbf{z}\mathbf{v}_\tau = (zV_\tau, zV_{\tau+1}, \dots, zV_{\tau+l-1}); \quad zV_s = \frac{V_s - \bar{V}_\tau}{\sigma_{V_\tau}}; \quad \tau \leq s \leq \tau + l - 1 \quad (3)$$

$$\bar{V}_\tau = \frac{1}{l} \sum_{s=\tau}^{\tau+l-1} V_s; \quad \sigma_{V_\tau} = \sqrt{\frac{1}{l-1} \sum_{s=\tau}^{\tau+l-1} (V_s - \bar{V}_\tau)^2}$$

where $1 \leq \tau \leq t - l + 1$, such that even when the scales of the asset values may be very different between windows, the patterns are scale free profiles (see Figure 1).

Lets look now at the set of the k l -dimensional behavior patterns that are most similar, as regards to their Euclidean distances, to the current pattern $\mathbf{z}\mathbf{v}_{t-l+1}$. Thus we need to identify k vectors that correspond to the smallest values of:

$$D_{t-l+1,\tau} = \|\mathbf{z}\mathbf{v}_{t-l+1} - \mathbf{z}\mathbf{v}_\tau\|$$

and place them in increasing order:

$$\mathbf{z}\mathbf{v}_{n_1(t,l)}, \mathbf{z}\mathbf{v}_{n_2(t,l)}, \dots, \mathbf{z}\mathbf{v}_{n_k(t,l)}$$

where $D_{t-l+1,n_1(t,l)} \leq D_{t-l+1,n_2(t,l)} \leq \dots \leq D_{t-l+1,n_k(t,l)}$.

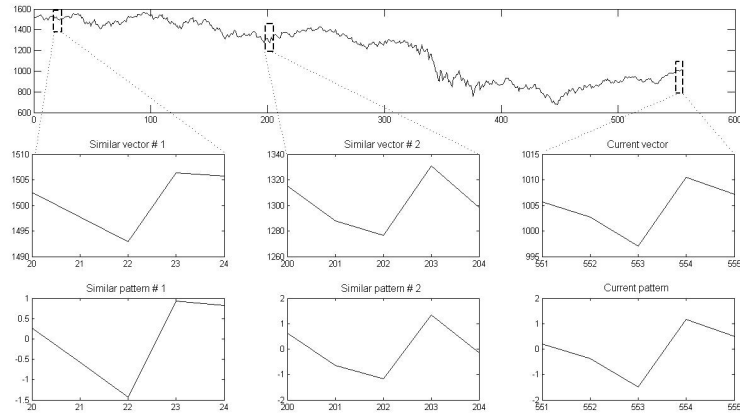


Figure 1: *An example of what results from the process of seeking patterns similar to the current pattern. Standardization permits us to compare each past pattern with the current pattern without taking into account the scale. Thus, what we use as inputs for the local GP are similar profiles.*

Now we construct, by appropriately scaling each $V_{n_i(t,l)+l}$ observed one step ahead of the i -th pattern, the respective standardized values of the asset which correspond to each of the patterns that define the neighborhood of the current pattern:

$$\mathbf{z}\mathbf{v}_{n_i(t,l)} \rightarrow zV_{n_i(t,l)+l} \equiv \frac{V_{n_i(t,l)+l} - \bar{V}_{n_i(t,l)}}{\sigma_{V_{n_i(t,l)}}}; \quad 1 \leq i \leq k \quad (4)$$

What we assume as being the information I_t , is the set of ordered pairs:

$$\mathbf{X}(t, l) \equiv \{\mathbf{x}_i(t, l), y_i(t, l); \quad 1 \leq i \leq k\}; \quad (5)$$

where

$$\mathbf{x}_i(t, l) \equiv \mathbf{z}\mathbf{v}_{n_i(t,l)}, \quad y_i(t, l) \equiv zV_{n_i(t,l)+l} \quad (6)$$

The second step in our method is to adjust the hyperparameters of the GP represented by a set of random variables $f_t(\mathbf{x}_i(t, l))$ and a noise term ϵ_i that satisfy the local condition:

$$y_i(t, l) = f_t(\mathbf{x}_i(t, l)) + \epsilon_i; \quad \epsilon_i \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \nu_t^2), \quad i = 1, 2, \dots, k \quad (7)$$

which we then use to generate the conditional predictive distribution

$$g(y^*(t, l) | \mathbf{X}(t, l), \mathbf{x}^*(t, l))$$

where $\mathbf{x}^*(t, l) \equiv \mathbf{z}\mathbf{v}_{t-l+1}$.

By following the standard procedure used in these cases and that is described in the relevant literature (see e.g. Brahim-Belhouari and A. Bermak (2004), Kim et al. (2005) and Rasmussen and Williams (2006)), we obtain that:

$$g(y^*(t, l) | \mathbf{X}(t, l), \mathbf{x}^*(t, l)) \propto \exp \left[-\frac{(y^*(t, l) - \bar{y}(t, l))^2}{2\hat{\sigma}^{*2}(t, l)} \right], \quad (8)$$

where $\bar{y}(t, l)$ can be identified as the expected value of $y^*(t, l)$ and $\hat{\sigma}^{*2}(t, l)$ with the variance, which are given respectively by:

$$\bar{y}(t, l) = \mathbf{c}(t, l) \mathbf{C}^{-1}(t, l) \mathbf{y}(t, l); \quad \hat{\sigma}^{*2}(t, l) = \gamma(t, l) - \mathbf{c}(t, l) \mathbf{C}^{-1}(t, l) \mathbf{c}^T(t, l) \quad (9)$$

with:

$$\mathbf{c}(t, l) = (c(\mathbf{x}^*(t, l), \mathbf{x}_1(t, l)), c(\mathbf{x}^*(t, l), \mathbf{x}_2(t, l)), \dots, c(\mathbf{x}^*(t, l), \mathbf{x}_k(t, l))), \quad (10)$$

$\mathbf{C}^{-1}(t, l)$ is the inverse matrix of:

$$\begin{pmatrix} c(\mathbf{x}_1(t, l), \mathbf{x}_1(t, l)) & c(\mathbf{x}_1(t, l), \mathbf{x}_2(t, l)) & \dots & c(\mathbf{x}_1(t, l), \mathbf{x}_k(t, l)) \\ c(\mathbf{x}_2(t, l), \mathbf{x}_1(t, l)) & c(\mathbf{x}_2(t, l), \mathbf{x}_2(t, l)) & \dots & c(\mathbf{x}_2(t, l), \mathbf{x}_k(t, l)) \\ \vdots & \vdots & \ddots & \vdots \\ c(\mathbf{x}_k(t, l), \mathbf{x}_1(t, l)) & c(\mathbf{x}_k(t, l), \mathbf{x}_2(t, l)) & \dots & c(\mathbf{x}_k(t, l), \mathbf{x}_k(t, l)) \end{pmatrix} \quad (11)$$

$$\mathbf{y}(t, l) = \begin{pmatrix} y_1(t, l) \\ y_2(t, l) \\ \vdots \\ y_k(t, l) \end{pmatrix}; \quad \gamma(t, l) = c(\mathbf{x}^*(t, l), \mathbf{x}^*(t, l)) \quad (12)$$

the superindex T denotes transposition, and $c(\mathbf{a}, \mathbf{b})$ is the covariance function defining the *GP*, here assumed as:

$$c(\mathbf{a}, \mathbf{b}) = \theta_t^2 \exp \left[-\frac{1}{2} (\mathbf{a} - \mathbf{b})^T M_t (\mathbf{a} - \mathbf{b}) \right] \quad (13)$$

where M_t is the diagonal matrix:

$$\begin{pmatrix} 1/s_{1t}^2 & \dots & 0 \\ \vdots & \ddots & 0 \\ 0 & 0 & 1/s_{dt}^2 \end{pmatrix}$$

In order to predict the asset value \hat{V}_{t+1} , we take the inverse of the standardization process on the expected value $\bar{y}(t, l)$:

$$\hat{V}_{t+1} = \sigma_{V_{t-l+1}} \cdot \bar{y}(t, l) + \bar{V}_{t-l+1}, \quad (14)$$

while the prediction for the uncertainty $\hat{\sigma}_{t+1}$ is:

$$\hat{\sigma}_{t+1} = \sigma_{V_{t-l+1}} \cdot \hat{\sigma}^* \quad (15)$$

It is worth emphasizing the local character of the proposed scheme, in the sense that in order to obtain the predictive distribution at $t + 1$ we use a *GP* whose hyperparameters $(\nu_t^2, \theta_t^2, s_{1t}, s_{2t}, \dots, s_{lt})$ are fitted to the information I_t . This itself is dependent on time, as indicated by Eq. 3-6, causing the *GP* parameters to change, providing a sufficiently flexible method for the modeling of non stationary series. As regards to the manner in which the parameters that intervene in the co-variance matrix are fixed, a criteria that is frequently used is to maximize the log-likelihood:

$$\mathcal{L} = -\frac{1}{2} \log[\det(\mathbf{C}(t, l))] - \frac{1}{2} \mathbf{y}(t, l)^T \mathbf{C}^{-1}(t, l) \mathbf{y}(t, l) - \frac{t}{2} \log 2\pi. \quad (16)$$

Furthermore, as indicated by the result of Eq. 8, when we apply models based on *GP* we obtain a predictive distribution, which makes it relatively simple to estimate the probability that a given event will occur. This fact is extremely interesting, from the financial point of view, in decision making. In particular, we show how models based on *GP* permit a straightforward evaluation of both *VaR* and *ES*.

We start by calculating the Value at Risk for a time horizon T and probability $\alpha \in (0, 1)$. The usual definition is:

$$VaR_{t+T}^\alpha = \sup\{v ; P[(R_{t+T} \leq v) | I_t] \leq \alpha\} \quad (17)$$

where $P(R|I_t)$ is the conditional distribution of the probabilities of the asset returns and I_t is, as before, the information that conditions what occurs at $t + T$. In other words, VaR is a simple way to estimate the minimal potential loss that, for a given time horizon, will be contained in the 100 α % of the worst case scenarios.

VaR may be estimated directly with the method we have described: we only have to calculate the return corresponding to the value, \hat{V}_{t+1}^α , below which the price of the asset may be found with probability α . The predictive distribution of the prices is given by the GP , thus \hat{V}_{t+1}^α satisfies the condition:

$$\alpha = \frac{1}{Z} \int_0^{\hat{V}_{t+1}^\alpha} \exp \left[-\frac{(x - \hat{V}_{t+1})^2}{2\hat{\sigma}_{t+1}^2} \right] dx \quad (18)$$

where Z is the normalization factor:

$$Z = \int_0^\infty \exp \left[-\frac{(x - \hat{V}_{t+1})^2}{2\hat{\sigma}_{t+1}^2} \right] dx = \sqrt{\frac{\pi}{2}} \hat{\sigma}_{t+1} \left[1 + \operatorname{erf} \left(\frac{\hat{V}_{t+1}}{\sqrt{2}\hat{\sigma}_{t+1}} \right) \right] \quad (19)$$

and $\operatorname{erf}(x)$ is the error function:

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-x^2) dx$$

Combining Eq. 18 and Eq. 19 we obtain:

$$\begin{aligned} \sqrt{\frac{\pi}{2}} \hat{\sigma}_{t+1} \left[\operatorname{erf} \left(\frac{\hat{V}_{t+1}}{\sqrt{2}\hat{\sigma}_{t+1}} \right) - \operatorname{erf} \left(\frac{\hat{V}_{t+1} - \hat{V}_{t+1}^\alpha}{\sqrt{2}\hat{\sigma}_{t+1}} \right) \right] &= \alpha Z = \\ &= \alpha \sqrt{\frac{\pi}{2}} \hat{\sigma}_{t+1} \left[1 + \operatorname{erf} \left(\frac{\hat{V}_{t+1}}{\sqrt{2}\hat{\sigma}_{t+1}} \right) \right] \end{aligned}$$

and therefore

$$\operatorname{erf} \left(\frac{\hat{V}_{t+1} - \hat{V}_{t+1}^\alpha}{\sqrt{2}\hat{\sigma}_{t+1}} \right) = (1 - \alpha) \operatorname{erf} \left(\frac{\hat{V}_{t+1}}{\sqrt{2}\hat{\sigma}_{t+1}} \right) - \alpha$$

or

$$\hat{V}_{t+1}^\alpha = \hat{V}_{t+1} - \sqrt{2}\hat{\sigma}_{t+1}erf^{-1}(\lambda_{t+1}) \quad (20)$$

with

$$\lambda_{t+1} = (1 - \alpha)erf\left(\frac{\hat{V}_{t+1}}{\sqrt{2}\hat{\sigma}_{t+1}}\right) - \alpha \quad (21)$$

According to Eq. 20, our estimation of VaR_{t+1}^α will be given by:

$$\widehat{VaR}_{t+1}^\alpha = \hat{R}_{t+1} - \frac{\sqrt{2}\hat{\sigma}_{t+1}}{V_t}erf^{-1}(\lambda_{t+1}) \quad (22)$$

where \hat{R}_{t+1} and $\hat{\sigma}_{t+1}$ are evaluated using Eq. 1, 14 and 15, and λ_{t+1} is given by Eq. 21.

We will now estimate ES with our model. If $\rho_{t+1}(R|I_t)$ is the conditional distribution of the returns of an asset and VaR_{t+1}^α the respective value at risk for a level α , then the expected shortfall, ES , is:

$$ES_{t+1}^\alpha = E(R_{t+1}|I_t, (VaR_{t+1}^\alpha > R_{t+1})) \propto \int_{-\infty}^{VaR_{t+1}^\alpha} R\rho_{t+1}(R|I_t)dR \quad (23)$$

such that, while VaR gives information about the least loss we could expect at a given confidence level; ES informs the worse expected loss that we can expect to occur.

In those cases where hedging needs to be designed, an adequate estimate of the ES is more informative than VaR . Lets take as a hypothetical situation that at a particular time t we know that the $VaR_{t+1}^{0.01}$ of an asset or a portfolio is -0.001 . There is thus a 99% probability that the return will exceed this value. However, the hedging plan will change if we estimate that the expected return for the remaining 1% is either -0.002 or -0.1 .

As for VaR , it is relatively simple to estimate ES using our GP based model. Effectively, since the predictive distributions of the prices are Gaussian, so are those of the arithmetic returns (given that these are obtained

from the prices by dividing them by yesterday's price). Thus, we can evaluate the previous integral by calculating the expected value $\widehat{ESV}_{t+1}^\alpha$ of the price of the asset, as long as this is below the $100\alpha\%$ quantile of the predictive distribution of the prices \hat{V}_{t+1}^α

$$\widehat{ESV}_{t+1}^\alpha \equiv \frac{1}{C_{t+1}^\alpha} \int_0^{\hat{V}_{t+1}^\alpha} x \exp \left[-\frac{(x - \hat{V}_{t+1})^2}{2\hat{\sigma}_{t+1}^2} \right] dx, \quad (24)$$

where C_{t+1}^α is the normalization factor:

$$\begin{aligned} C_{t+1}^\alpha &= \int_0^{\hat{V}_{t+1}^\alpha} \exp \left[-\frac{(x - \hat{V}_{t+1})^2}{2\hat{\sigma}_{t+1}^2} \right] dx = \\ &= \sqrt{\frac{\pi}{2}} \hat{\sigma}_{t+1} \left[\operatorname{erf} \left(\frac{\hat{V}_{t+1}}{\sqrt{2}\hat{\sigma}_{t+1}} \right) - \operatorname{erf} \left(\frac{\hat{V}_{t+1} - \hat{V}_{t+1}^\alpha}{\sqrt{2}\hat{\sigma}_{t+1}} \right) \right], \end{aligned}$$

The integral in Eq. 24 can be readily evaluated:

$$\widehat{ESV}_{t+1}^\alpha = \hat{V}_{t+1} - \sqrt{\frac{2}{\pi}} \hat{\sigma}_{t+1} \left\{ \frac{\left[\exp \left(-\frac{(\hat{V}_{t+1}^\alpha - \hat{V}_{t+1})^2}{2\hat{\sigma}_{t+1}^2} \right) - \exp \left(-\frac{\hat{V}_{t+1}^2}{2\hat{\sigma}_{t+1}^2} \right) \right]}{\left[\operatorname{erf} \left(\frac{\hat{V}_{t+1}}{\sqrt{2}\hat{\sigma}_{t+1}} \right) - \operatorname{erf} \left(\frac{\hat{V}_{t+1} - \hat{V}_{t+1}^\alpha}{\sqrt{2}\hat{\sigma}_{t+1}} \right) \right]} \right\}$$

and from this result we get our estimation for $\widehat{ES}_{t+1}^\alpha$:

$$\begin{aligned} \widehat{ES}_{t+1}^\alpha &= \frac{\widehat{ESV}_{t+1}^\alpha}{V_t} - 1 = \hat{R}_{t+1} - \\ &- \sqrt{\frac{2}{\pi}} \frac{\hat{\sigma}_{t+1}}{V_t} \left\{ \frac{\left[\exp \left(-\frac{(\hat{V}_{t+1}^\alpha - \hat{V}_{t+1})^2}{2\hat{\sigma}_{t+1}^2} \right) - \exp \left(-\frac{\hat{V}_{t+1}^2}{2\hat{\sigma}_{t+1}^2} \right) \right]}{\left[\operatorname{erf} \left(\frac{\hat{V}_{t+1}}{\sqrt{2}\hat{\sigma}_{t+1}} \right) - \operatorname{erf} \left(\frac{\hat{V}_{t+1} - \hat{V}_{t+1}^\alpha}{\sqrt{2}\hat{\sigma}_{t+1}} \right) \right]} \right\} \quad (25) \end{aligned}$$

In the next section we evaluate the performance of the model by showing the results obtained by calculating $\widehat{VaR}_{t+1}^{0.01}$ following Eq. 22 and $\widehat{ES}_{t+1}^{0.01}$ according to Eq. 25, for a series taken from the S&P500 index.

3. Backtesting VaR and ES estimations

The set of recommendations established by the Basel Committee is currently used to measure the quality of VaR estimates. These are statistical validation measurements (backtesting) based on the quantification of exceptions:

$$e_{t+1}^{\alpha} = \Theta(\widehat{VaR}_{t+1}^{\alpha} - R_{t+1})$$

that can occur when estimating VaR_{t+1}^{α} , where $\Theta(x)$ is Heavisides step function ($\Theta(x) = 1(0)$ when $x > 0$ (in any other case)).

Whichever the method used for calculating VaR , the condition that must be satisfied is:

$$E(e_{t+1}^{\alpha} | I_t) = \alpha$$

and what is examined are the different consequences of this condition.

Probably, the most utilized format is that if there are a total of n calculated values, the total number of exceptions:

$$x = \sum_{t=0}^{n-1} e_{t+1}^{\alpha}$$

follows a binomial distribution with parameters n and α . Thus for a 5% confidence level for $n = 250$ evaluations (that for daily data represents approximately one year) and a value of $\alpha = 0.01$, the method will only be rejected when $x > 5$.

For the example of the series taken from the S&P500 index represented in Figure 2, we can use Eq. 22 to calculate $\widehat{Var}_{t+1}^{0.01}$ over three consecutive windows of $n = 250$ data points (i.e. approximately three continuous years) and not once do we obtain more exceptions than those considered acceptable according to the validation criteria (3, 5 and 5 exceptions, respectively).

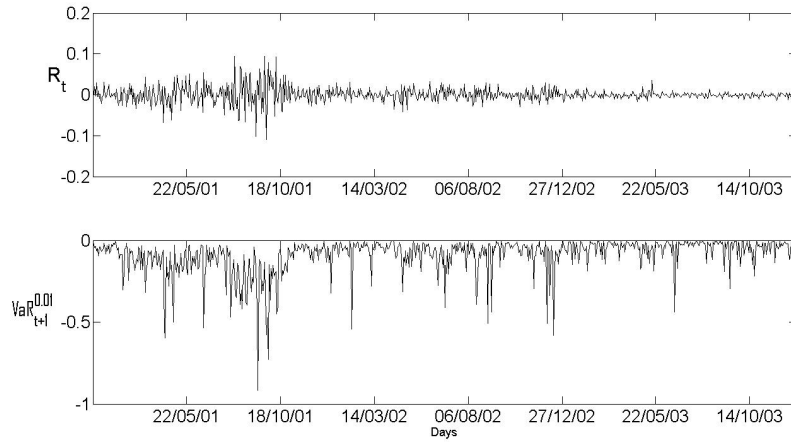


Figure 2: *In the upper part of the figure a segment representing the daily returns of the S&P500 index over a period of three years is shown. Observe the period of high volatility characterized by large fluctuations in the returns. The inferior part of the figure shows our $VaR_{t+1}^{0.01}$ estimates.*

As regards the evaluation of ES estimates done with the model, for each of the exceptions (total of 13) we compared the ES that we get using our model with the observed returns. Specifically, if we use $t_i^e, i = 1, \dots, 13$ to denote the 13 values of the times that correspond to the exceptions, we can calculate their mean squared error:

$$\frac{1}{\rho_e} \sqrt{\sum_{i=1}^{13} \left(\widehat{ES}_{t_i^e} - R_{t_i^e} \right)^2}, \quad \rho_e = \sqrt{\sum_{i=1}^{13} \left(R_{t_i^e} - \frac{1}{13} \sum_{j=1}^{13} R_{t_j^e} \right)^2}$$

This gives an error of 0.12, indicating the quality of the fit achieved by our model.

4. Discussion

There are several reasons why the behavior of financial markets is difficult to model, above all, because each agent involved prefers that others do not have access to a model that permits them to correctly anticipate fluctuations in the market. If they did, and could thus exchange uncertainty for predictability, circumstantial informational advantages that could be used to increase net worth would be lost, which at the end of the day is what this is all about.

Apart from practical motivations whose importance is obvious, the difficulties inherent to modeling combined with the ever increasing possibility of having access to large volumes of data and computer software, have stimulated the development of empirical models that, as well as predicting what the system will or will not do, also attempt to estimate the degree of uncertainty that this may present.

Models based on Gaussian processes incorporate this option in a direct way as a consequence of their structure. This is advantageous due to the simplification of the theory involved and also from a practical point of view. They do this by assuming that the series of values of an asset is a realization of a random process and that this is Gaussian. Thus, predictive conditional distributions that are Gaussian centered on the expected values of the asset are obtained, whose standard deviations fix the respective volatility values and whose quantiles correspond to estimates of risk indexes such as *VaR* and *ES*. All this without the necessity of introducing additional assumptions, such as for example, the evolution of the conditional variance and/or the distribution of the observed returns, which are necessary in ARCH/GARCH type models and in models of stochastic volatility.

Another beneficial aspect is the ease with which we can include the assumptions as relevant information that condition the predictive distribution in each case. One of these benefits can be when we suppose that the information is given by the way in which the system evolves from similar initial situations. The problem is thus reduced to specifying the set of vectors that this information represents and to evaluate the corresponding covariance function.

With regards this last point, it is worth mentioning that even though all the results we have presented correspond to a particular selection of the covariance function, the key issue is that whatever the proposed function is, this generates a covariance matrix whose elements are much higher among vectors that are similar than among those that are not, which falls rapidly with the distance between vectors. Finally, the models presented here permit

the establishment of simple criteria for risk evaluation that could be useful in the design of dynamic portfolios with optimal risk, which could in turn result in portfolios that generate higher equity than that produced by minimum risk portfolios. These are two of the aspects we wish to focus on in a further study.

5. References

References

Basel Committee on Banking Supervision, 1995. An Internal Model-Based Approach to Market Risk Capital Requirements, BIS, Basel, Switzerland.

Basel Committee on Banking Supervision, 1996. Supervisory Framework for the Use of Backtesting in Conjunction with the Internal Model-Based Approach to Market Risk Capital Requirements, BIS, Basel, Switzerland.

S. Brahim-Belhouari and A. Bermak, 2004. Gaussian process for nonstationary time series prediction. *Computational Statistics & Data Analysis*, **47**, pp. 705–712.

S.T.S. Bukkapatnam and C. Cheng, 2010. Forecasting the evolution of nonlinear and nonstationary systems using recurrence-based local Gaussian processes models. *Phy. Rev. E*, **82**, doi: 10.1103/PhysRevE.82.056206 .

M. Deisenroth, C. Rasmussen and J. Peters, 2009. Gaussian process dynamic programming. *Neurocomputing*, **72**(7-9), pp. 1508-1524.

F. di Sciascio and A. Amicarelli, 2008. Biomass estimation in batch biotechnological processes by bayesian gaussian process regression. *Computers and Chemical Engineering*, **32**(12), pp. 3264-3273.

- R. Engle, S. Focardi, and F. Fabozzi, 2008. Handbook Series in Finance, chapter ARCH-GARCH Models in Applied Financial Econometrics. John Wiley & Sons, doi: 10.1002/9780470404324.hof003060.
- L. Gosoniua, P. Vounatsoua, N. Sogobab, N. Mairea & T. Smitha, 2009. Mapping malaria risk in West Africa using a Bayesia nonparametric non-stationary model. *Computational Statistics & Data Analysis*, **53**, Issue 9, pp. 3358-3371.
- M.T. Haavelmo, 1944. The probability approach in econometrics. *Econometrica*, **12**, pp. 1-15.
- J. Hull, 2007. Risk Management and Financial Institutions, Prentice Hall.
- H.M. Kim , B. K. Mallick and C.C. Holmes, 2005. Analyzing Nonstationary Spatial Data Using Piecewise Gaussian Processes. *Journal of the American Statistical Association*, **100**, No. 470, pp. 653-668.
- J. Ko and D. Fox, 2009. Gp-bayesfilters: Bayesian filtering using gaussian process prediction and observation models. *Autonomous Robots*, **27**(1), pp. 75-90.
- C.E. Rasmussen & C.K.I. Williams, 2006. Gaussian Process for Machine Learning. MIT Press.
- J. Wang, D. Fleet, and A. Hertzmann, 2008. Gaussian process dynamical models for human motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **30**(2), pp. 283-298.

6. Acknowledgement

This paper is part of the PhD Thesis of I. García at Universidad Simón Bolívar, which was financed through a fellowship by the Academia Nacional de Ciencias Físicas, Matemáticas y Naturales of Venezuela. Venezuela.